Ensemble Spread as a Precursor for Extreme Space Weather Events

Erin Lynch

Surja Sharma

Eugenia Kalnay

Kayo Ide

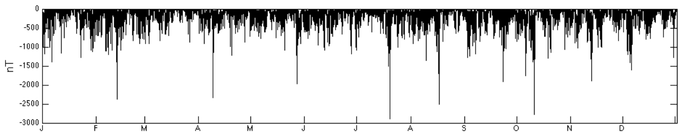
Data Assimilation for Data-Derived Models

- Motivation
  - Data assimilation is a powerful tool used to improve forecasts of nonlinear systems
  - Data assimilation produces more accurate initial conditions by combining estimates from previous model forecasts and observations
  - Ensemble techniques provide an efficient way to estimate model error by sampling nearby trajectories
- Goal:
  - We apply the Ensemble Transform Kalman Filter (ETKF) to a data-derived model to produce better forecasts
  - The spread of the forecast ensembles gives information about instability and can be used to forecast extreme events

Phase space model constructed from a time series of scalar observations

Time Delay Vectors create multivariate phase space vectors from scalar time series
- Preserves features of the attractor
- Parameters to tune: time delay and dimension

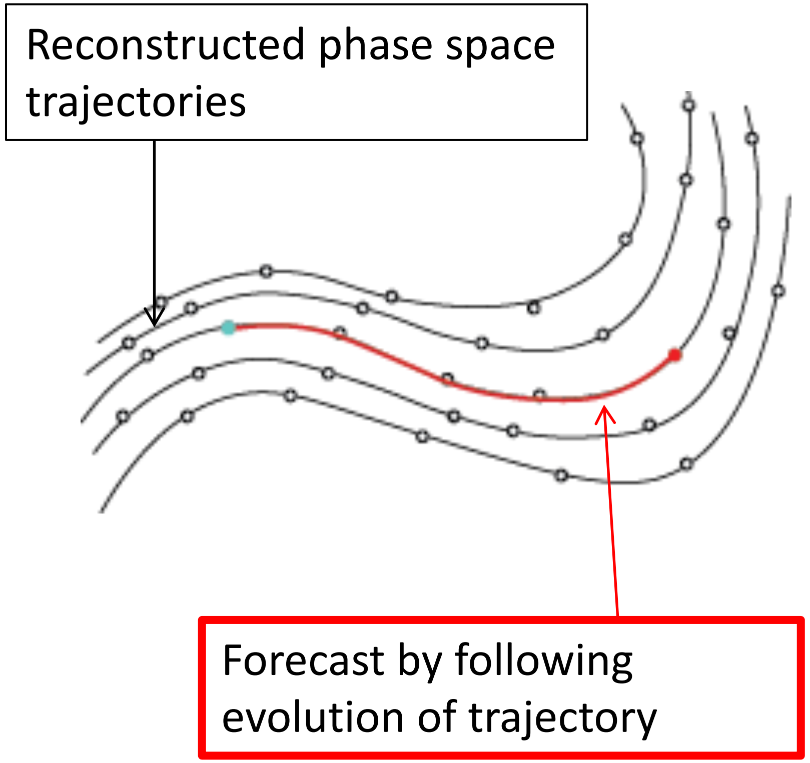
$$\{x(t_1), x(t_2), \ldots, x(t_N)\}$$



$$\begin{pmatrix} x(t) \\ x(t+\tau) \\ x(t+2\tau) \\ \vdots \\ x(t+(m-1)\tau) \end{pmatrix}$$

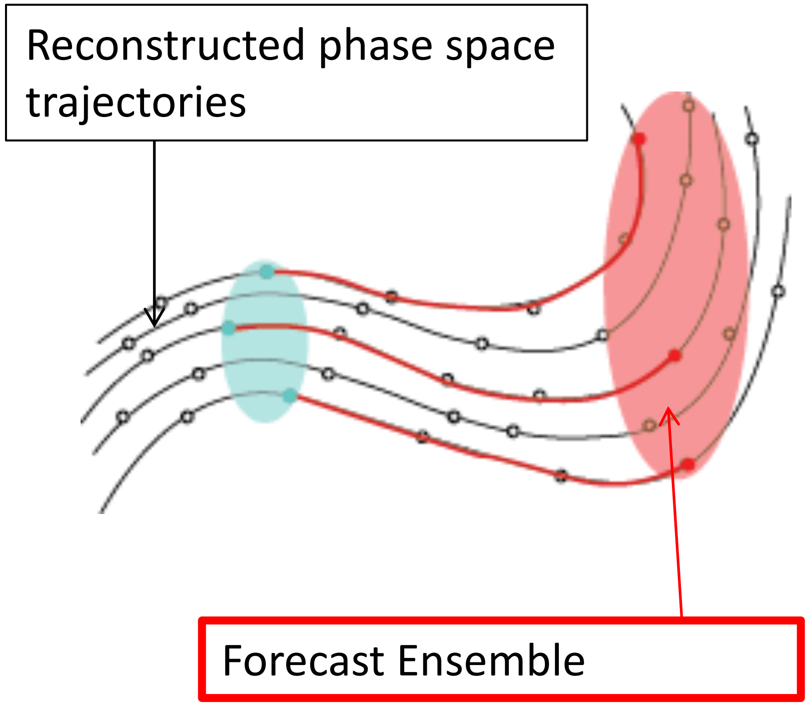
Singular Spectrum Analysis reduces the dimension of the phase space
- Identify modes of variability
- Keep modes that represent most of the signal variance
- Reject those that correspond to noise

Forecasts using a reconstructed phase space model



- Forecasts are made by following the reconstructed phase space trajectories
- Ensemble of forecasts made from slightly perturbed initial conditions

Forecasts using a reconstructed phase space model

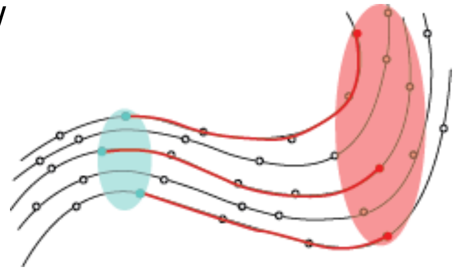


- Forecasts are made by following the reconstructed phase space trajectories
- Ensemble of forecasts made from slightly perturbed initial conditions

Ensemble Transform Kalman Filter (ETKF) with a Data-Derived Model

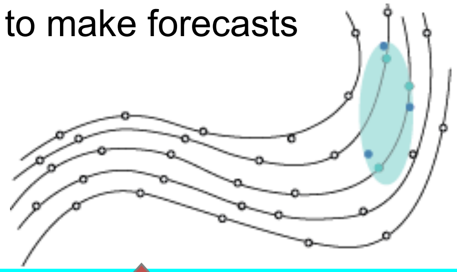
“Model” Forecast

Use a dense data set of points on the attractor (model) to advance NN analysis ensemble to the end of the analysis window



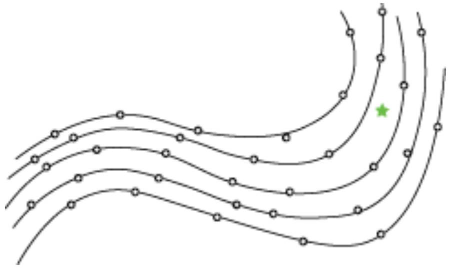
Nearest Neighbors (NN)

Locate nearest neighbors of analysis ensemble members to serve as analogs to make forecasts



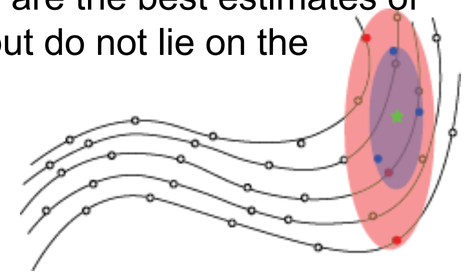
Observations

Observations of a single variable (i.e. the AL index) become multivariate when embedded

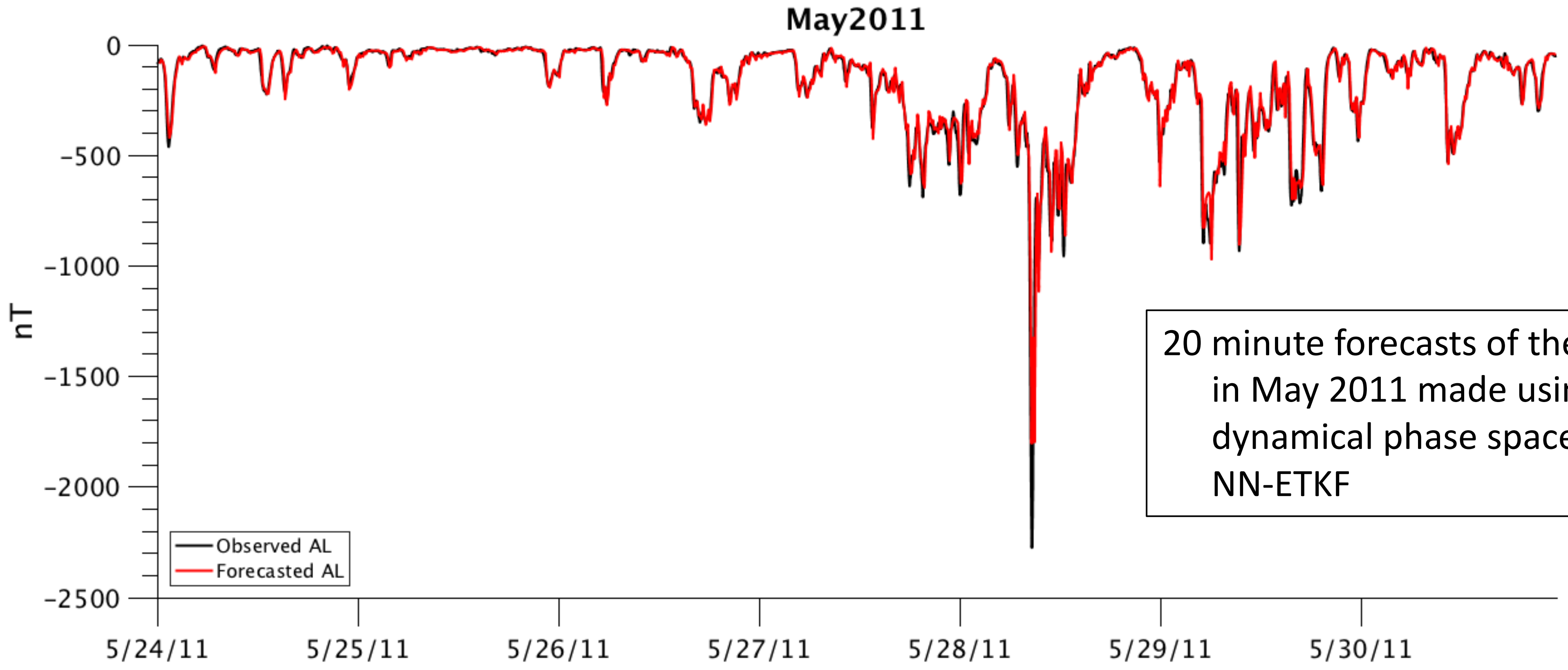


Analysis

Analysis ensemble members computed using the ETKF are the best estimates of the true state, but do not lie on the attractor

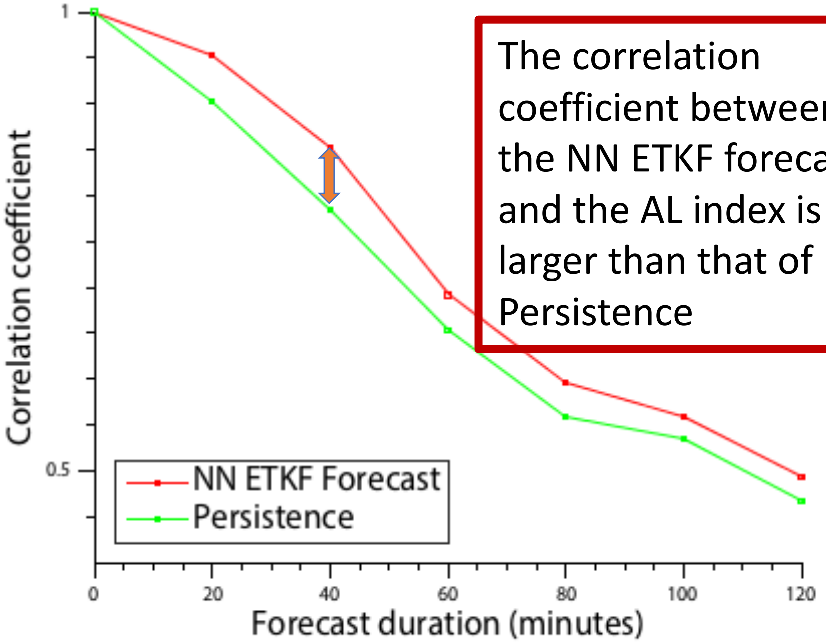


Forecast results using ETKF and date-derived model



20 minute forecasts of the HSS event in May 2011 made using the dynamical phase space model and NN-ETKF

Forecasts improved by ETKF



The correlation coefficient between the NN ETKF forecast and the AL index is larger than that of Persistence

Skill Score

$$SS = 1 - \frac{MSE_{\text{NN-ETKF}}}{MSE_{\text{Persistence}}}$$

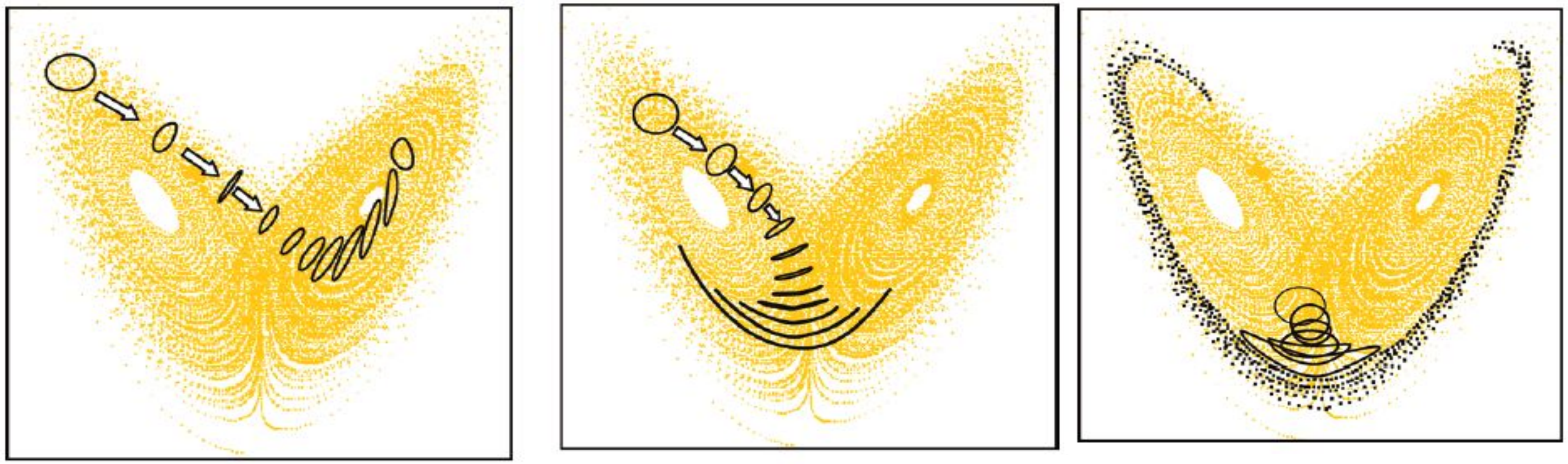
For 20 minute forecasts

$$SS = 0.58$$

Positive skill scores implies NN-ETKF forecast is performing better than persistence

Ensemble spread as a precursor to extreme events

- Consider the transition from one wing of Lorenz attractor to the other an extreme event

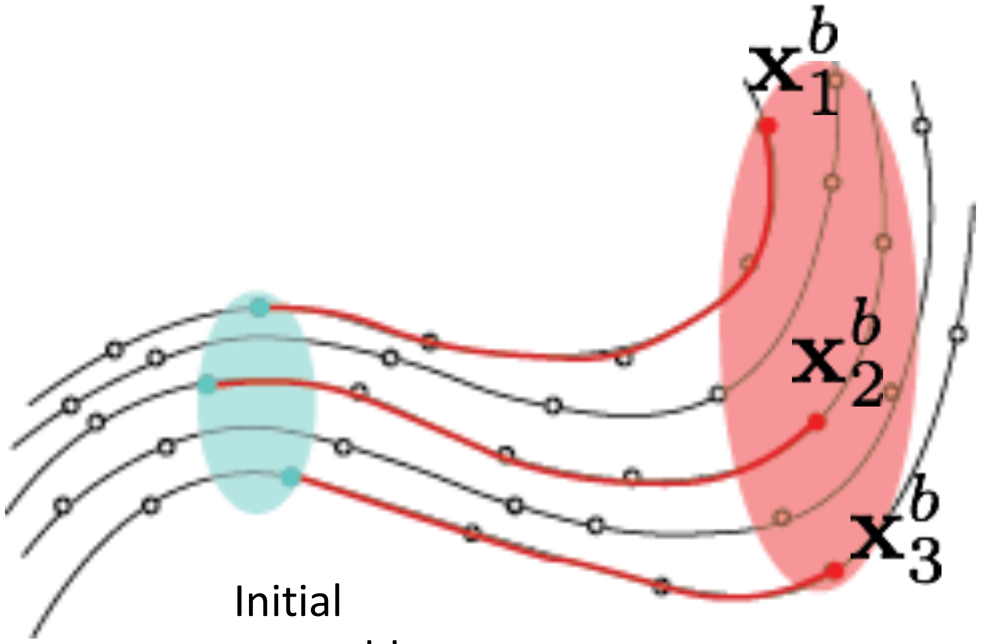


Non-extreme event

Extreme event

Figure from Palmer et al. 1999

Ensemble Spread



Initial ensemble

Forecast Ensemble

Forecast Ensemble

$$\mathbf{X}^b = \{\mathbf{x}_1^b, \ldots, \mathbf{x}_M^b\}$$

Ensemble mean

$$\bar{\mathbf{x}}^b = M^{-1} \sum_{i=1}^{M} \mathbf{x}_i^b$$

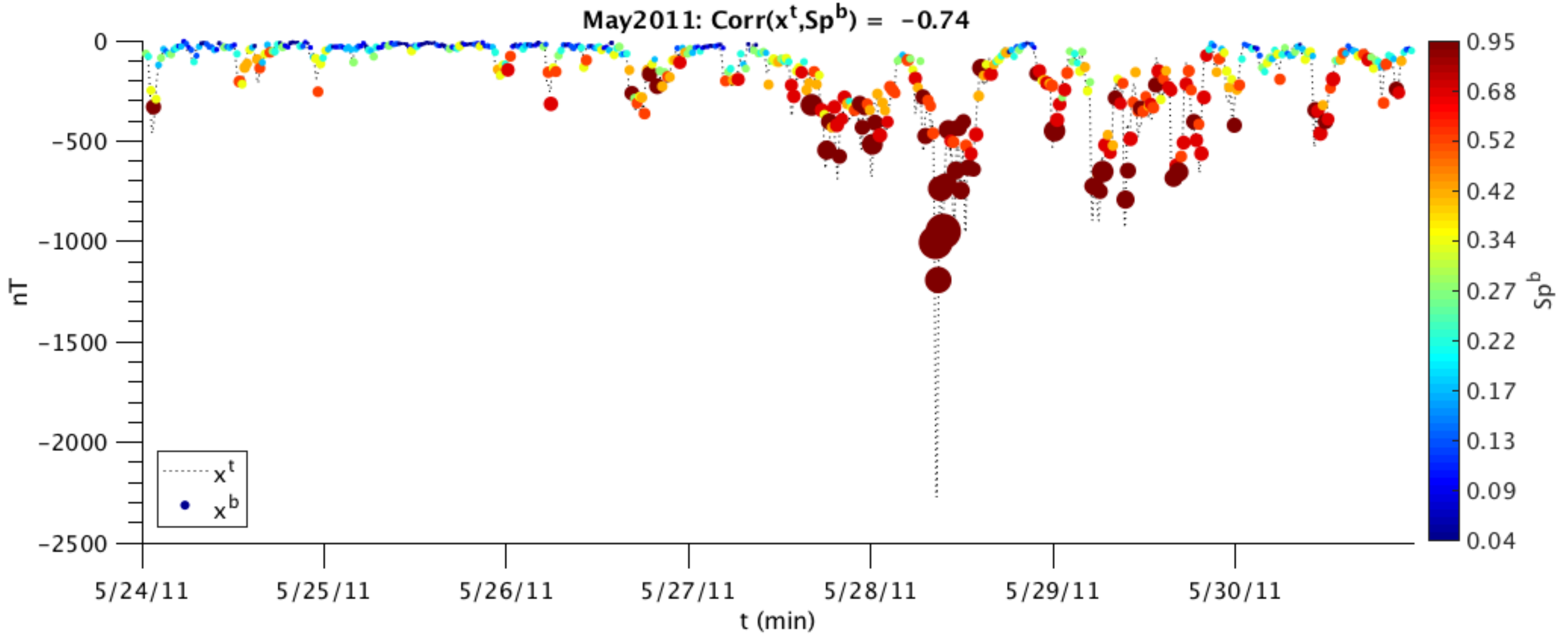
Ensemble spread

$$\hat{\mathbf{X}}^b = \{\mathbf{x}_1^b - \bar{\mathbf{x}}^b, \ldots, \mathbf{x}_M^b - \bar{\mathbf{x}}^b\}$$

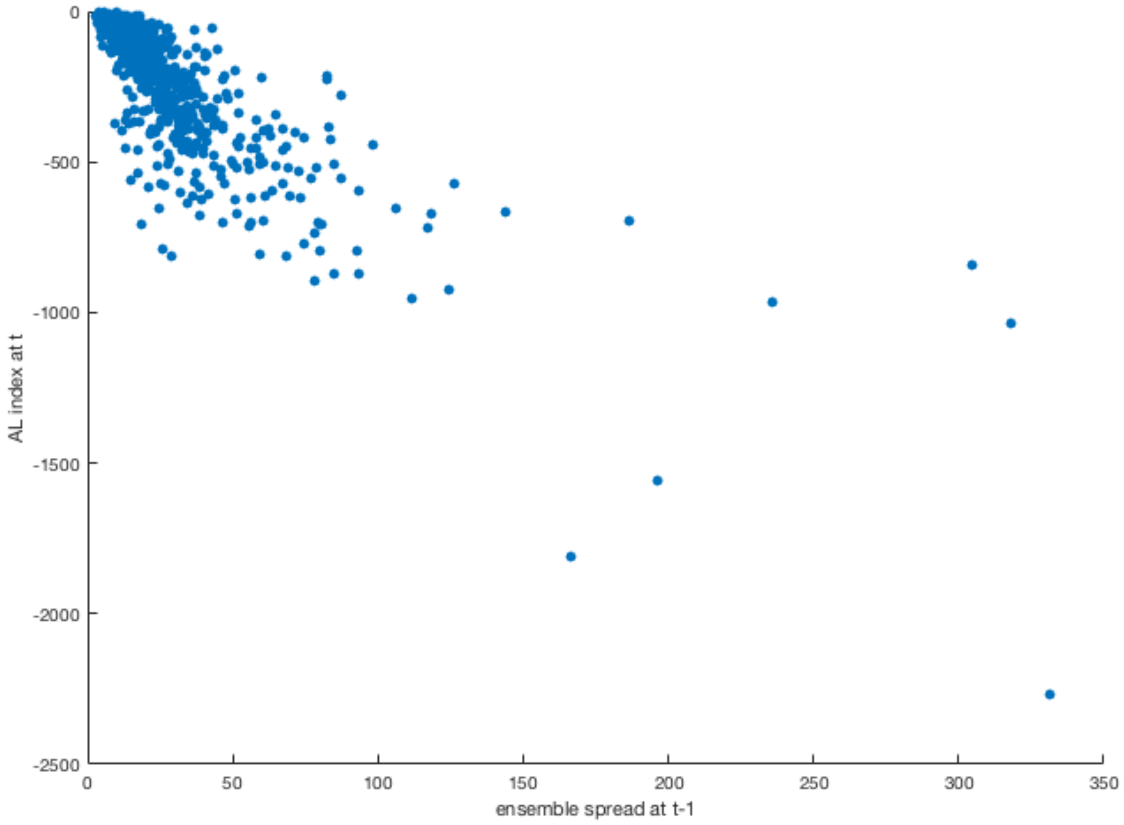
$$\mathbf{P}^b = (M-1)^{-1} \hat{\mathbf{X}}^b (\hat{\mathbf{X}}^b)^T$$

$$\mathrm{Sp}^b = \sqrt{\mathrm{Tr}(\mathbf{P}^b)}$$

Ensemble Spread Correlates with the Magnitude of AL



Ensemble spread time lagged correlation with the value of the AL index

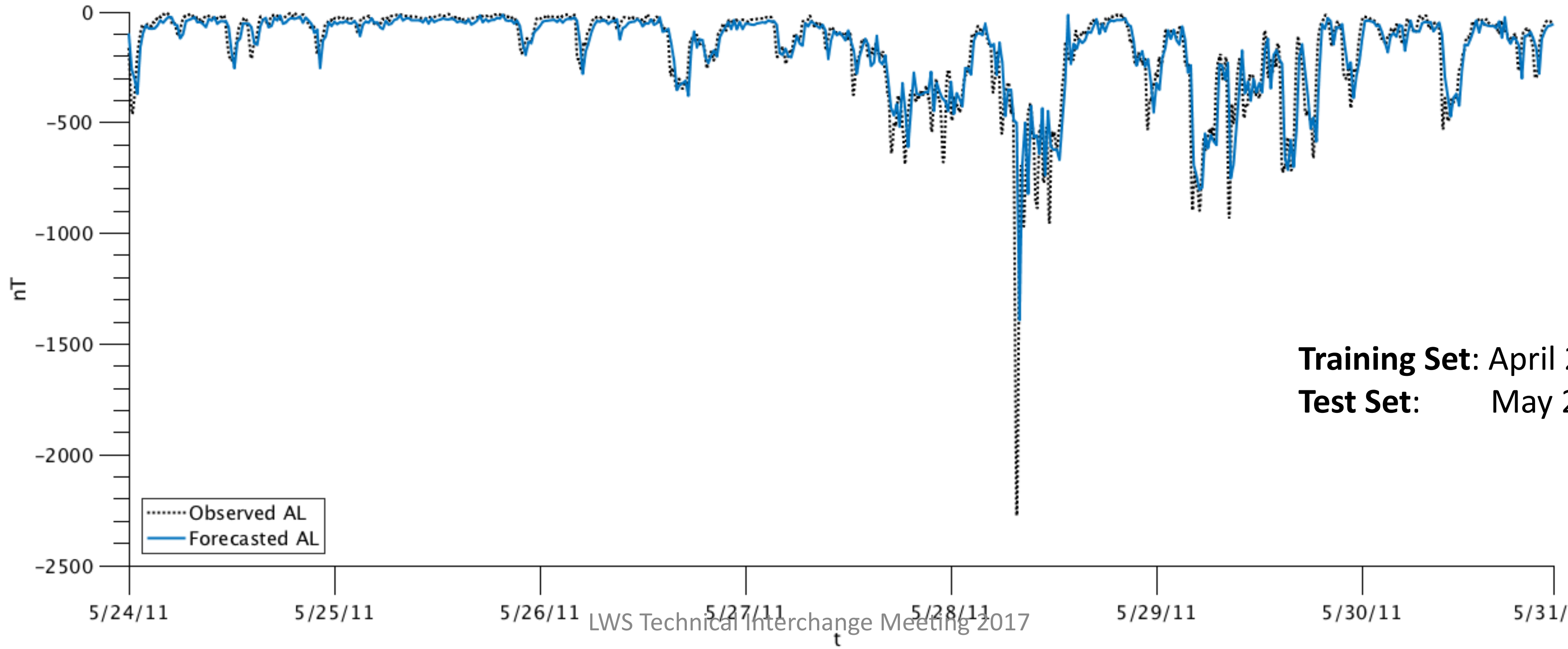


- The correlation between the spread of a forecast and the value of the AL index 20 minutes later is high

- Indicating that the spread of the previous forecast is a good predictor of the future value of the AL index

Ensemble spread to forecast $\{x^b, x^p, S^b, \frac{dS^b}{dt}\}$

$$AL(t+1) = b_0 + b_1 \bar{x}^b(t) + b_2 AL(t-1) + b_3 S^b(t) + b_4 S^b(t-1)$$



Training Set: April 2011
Test Set: May 2011 event

Summary

- We have successfully applied data assimilation to a data-derived model to produce
- Ensemble forecasts using the ETKF improve predictions of the AL index
- We are able to identify the ensemble spread as an indicator of extreme events and can use as a precursor to predict their onset